







SACRO: Semi-Automated Checking of Research Outputs

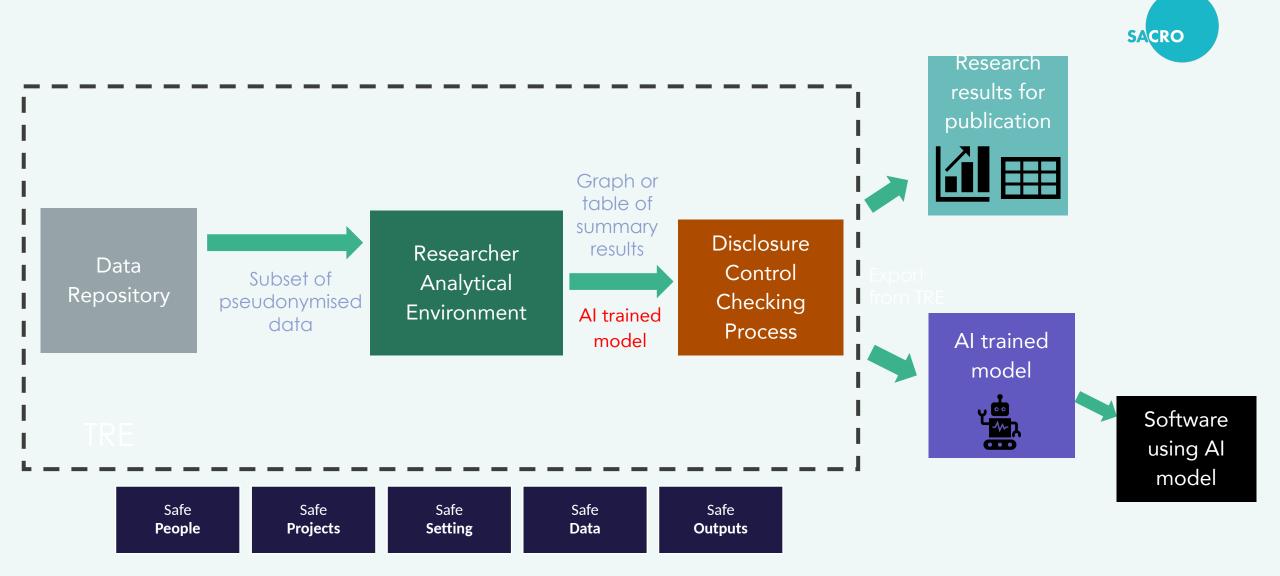
Professor Jim Smith,

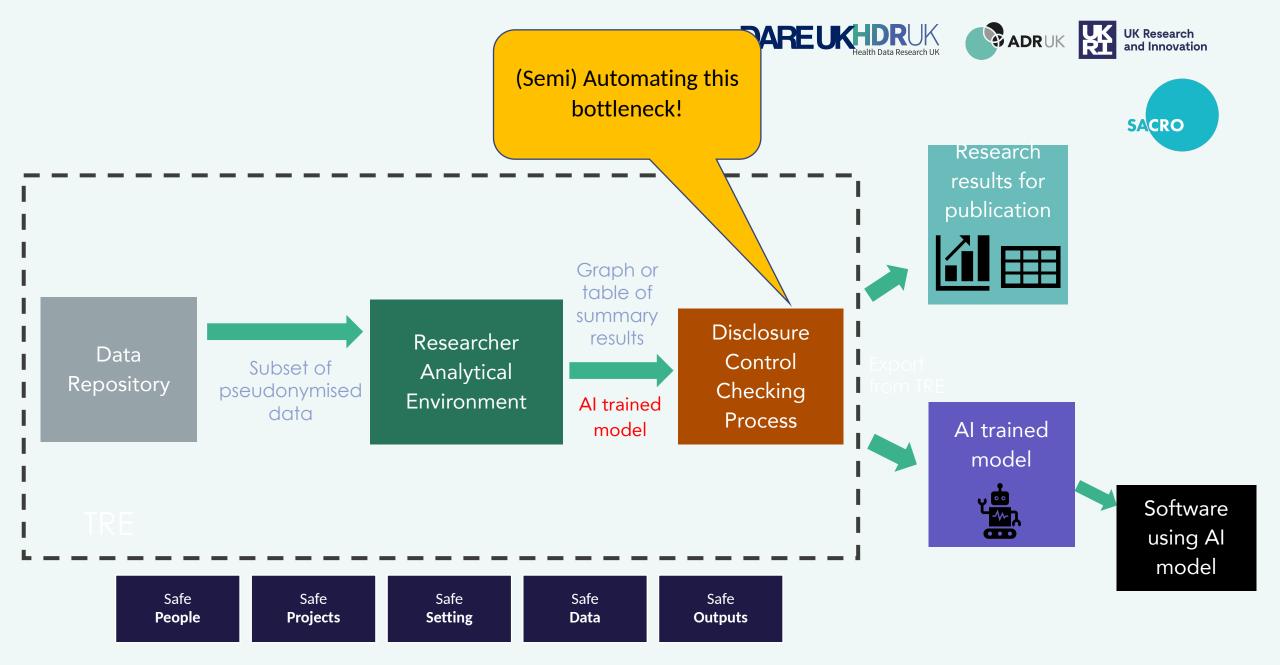
University of the West of England











SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

















Universities

- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE









Universities

- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE

Public Data Bodies

- Health Data Research UK
- NHS Scotland
- Public Health Scotland
- Research Data Scotland









Universities

- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE

Public Data Bodies

- Health Data Research UK
- NHS Scotland
- Public Health Scotland
- Research Data Scotland

TRES

- DASH (Aberdeen/Grampian)
- DataLoch (Edinburgh)
- HIC (Dundee)
- eDRIS (Public Health Scot)
- OpenSafely (Oxford)









Universities

- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE

Public Data Bodies

- Health Data Research UK
- NHS Scotland
- Public Health Scotland
- Research Data Scotland

TRES

- DASH (Aberdeen/Grampian)
- DataLoch (Edinburgh)
- HIC (Dundee)
- eDRIS (Public Health Scot)
- OpenSafely (Oxford)

External steering group:

UK: Office for National Statistics, ESRC, DARE Global: Eurostat, SDC-GESIS, ICPSR (US)



























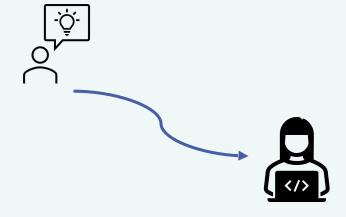












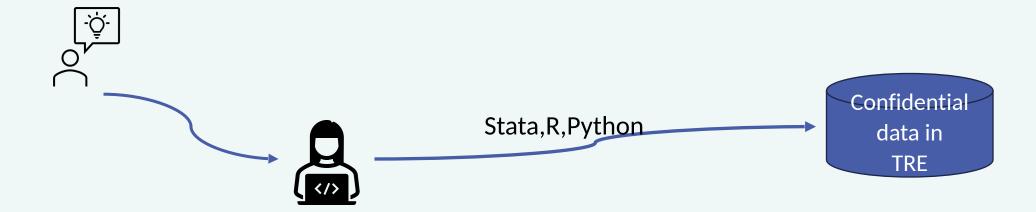










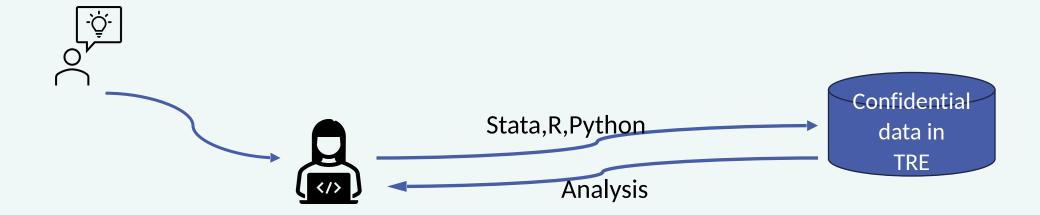










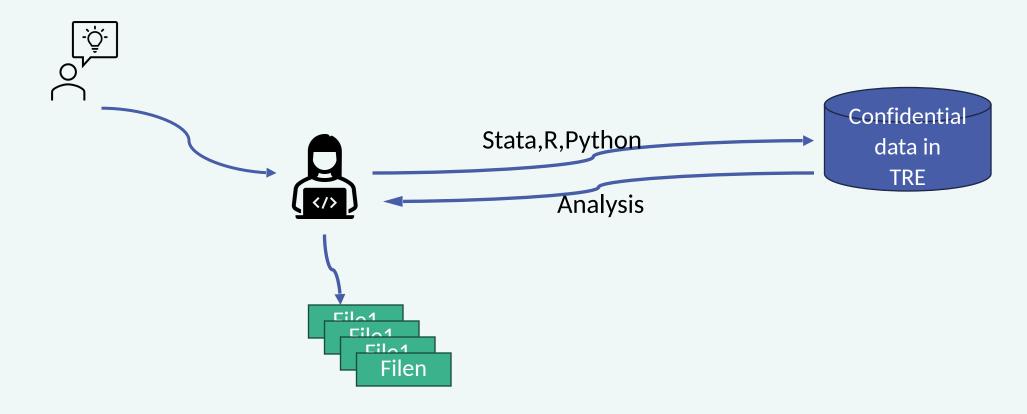










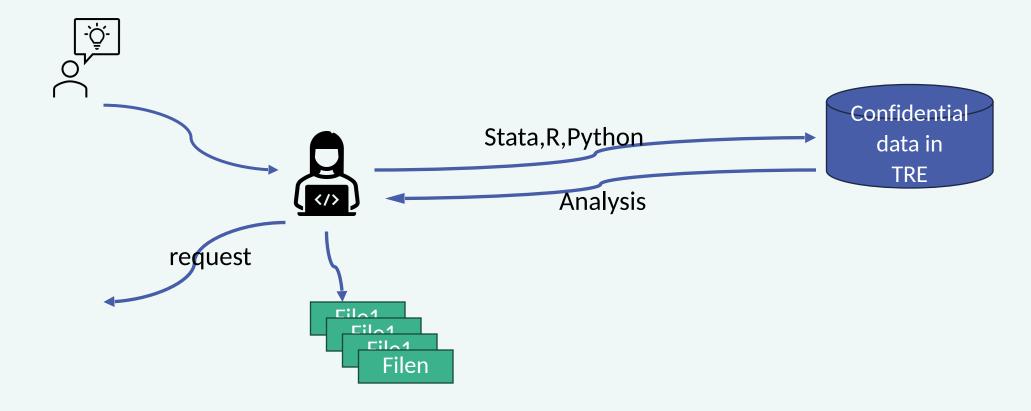










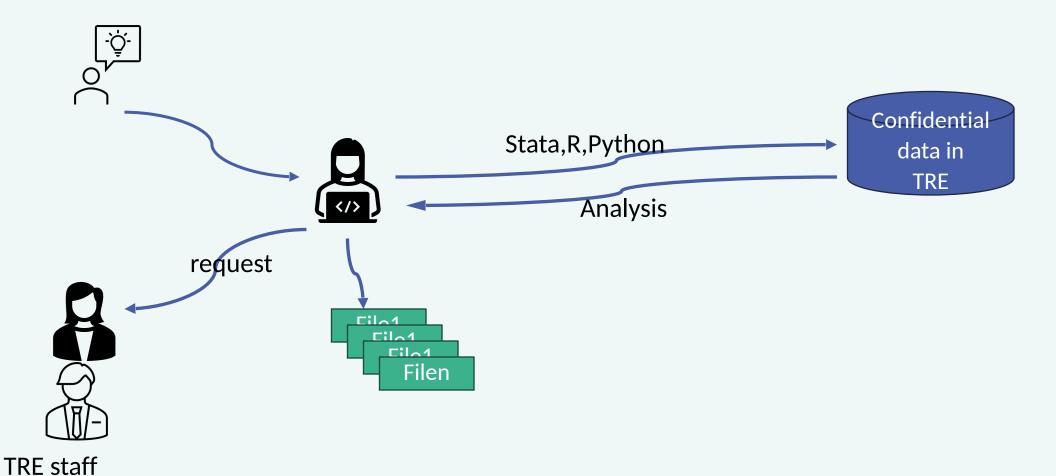










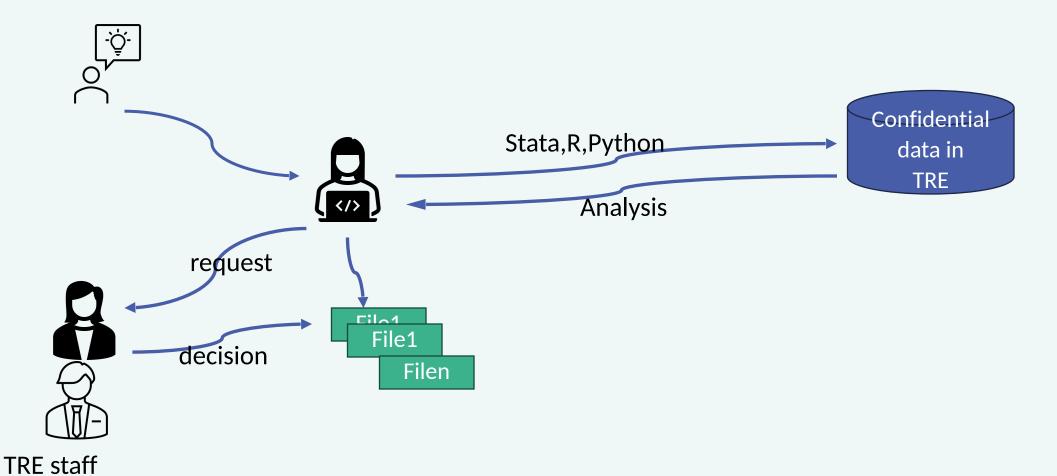


















































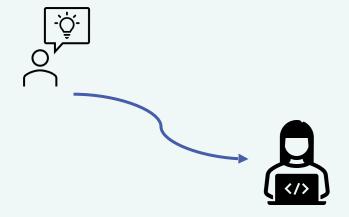














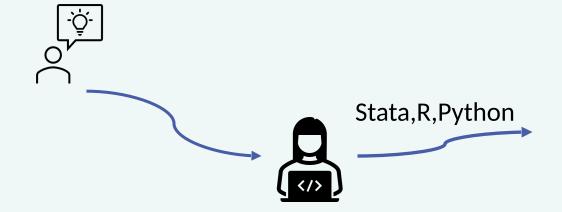






SACRO

SACRO in a nutshell



Confidential data in TRE

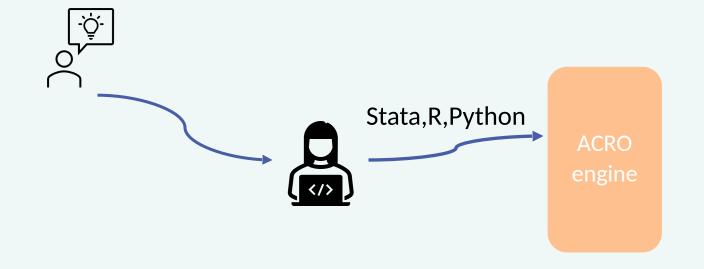






SACRO

SACRO in a nutshell



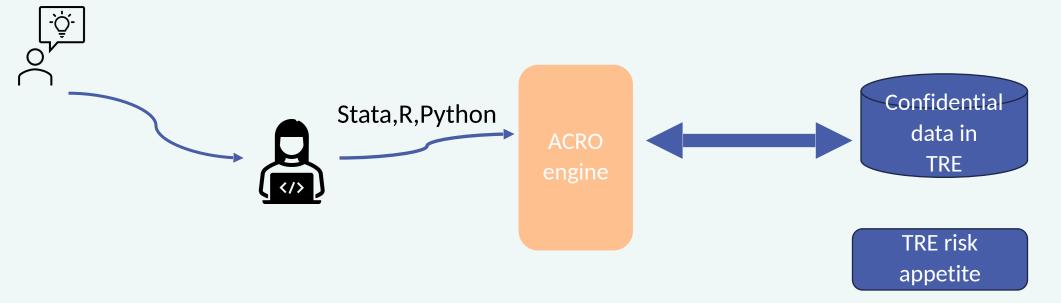
Confidential data in TRE









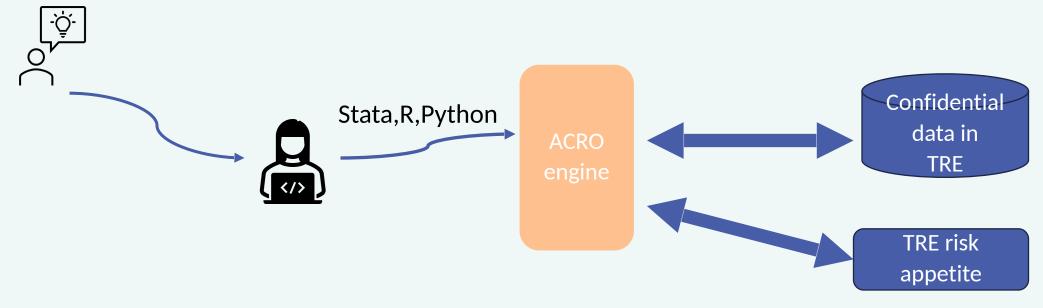










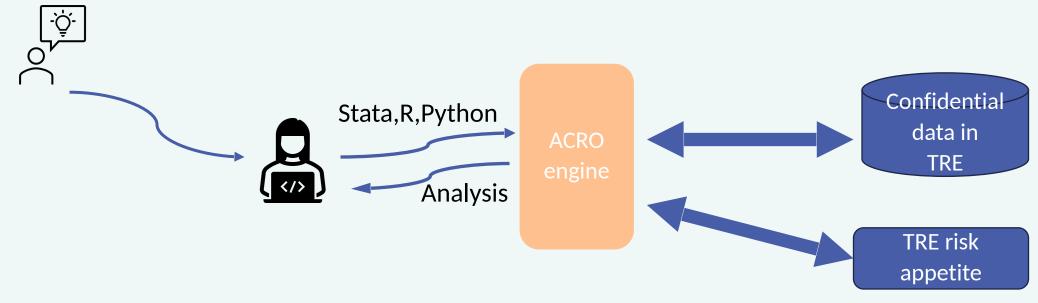










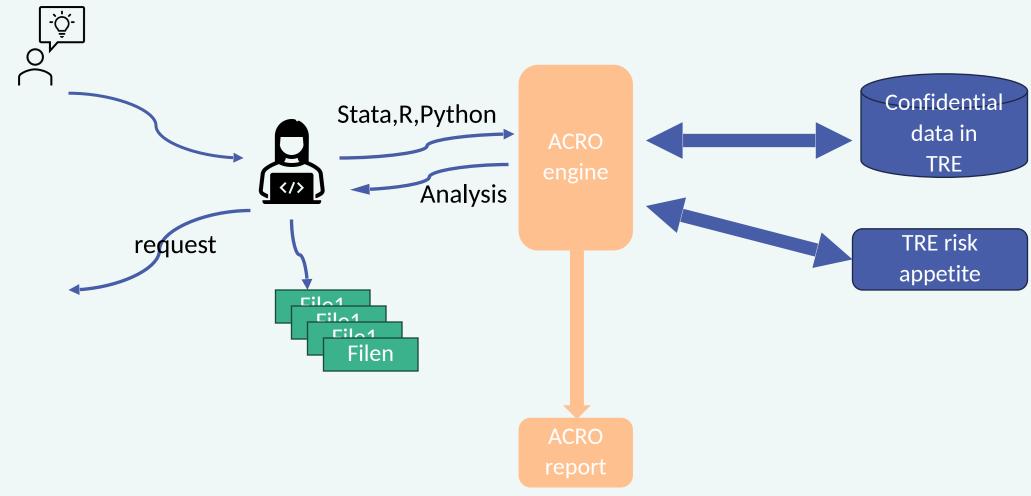










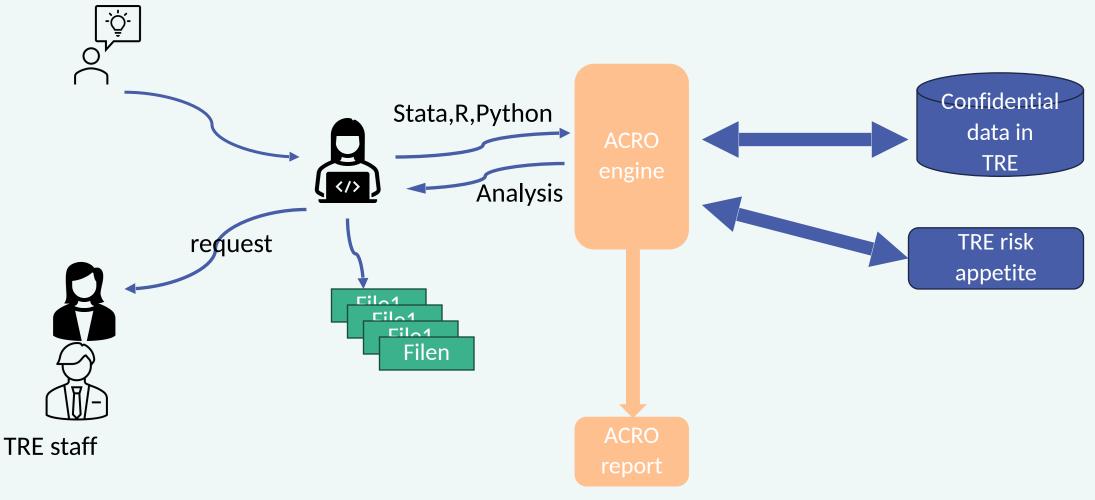










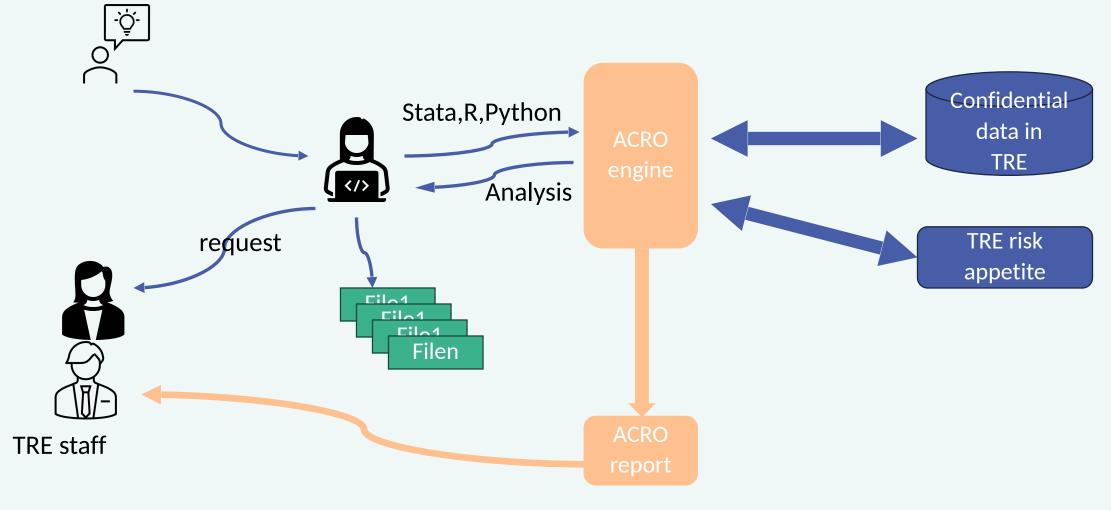










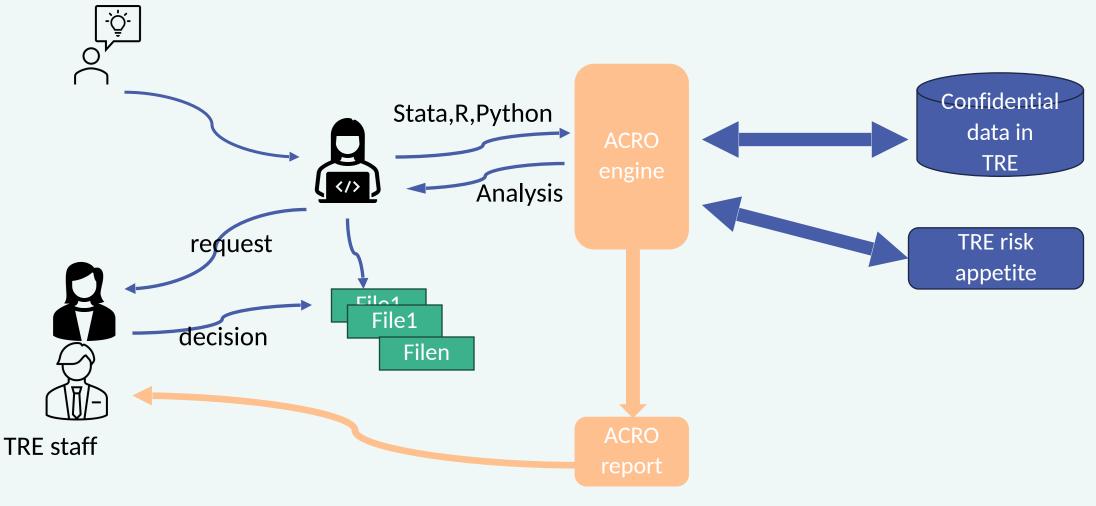




















SACRO

Except that we:











Except that we:

Run a range of "inference" attacks"











Except that we:

- Run a range of "inference" attacks"
- Aim to support more 'user journeys'











Except that we:

- Run a range of "inference" attacks"
- Aim to support more 'user journeys'
- Don't have a set of 'tried and trusted' guidelines to work with



















ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- acro.crosstab()
- acro.pivot_table()

Using the same api as statsmodels, adding checks for DoF:









ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- acro.crosstab()
- acro.pivot_table()

Using the same api as statsmodels, adding checks for DoF:

acro.logit(), acro.logitr()









ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- acro.crosstab()
- acro.pivot_table()

Using the same api as statsmodels, adding checks for DoF:

- acro.logit(), acro.logitr()
- acro.ols(), acro.olsr()









ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- acro.crosstab()
- acro.pivot_table()

Using the same api as statsmodels, adding checks for DoF:

- acro.logit(), acro.logitr()
- acro.ols(), acro.olsr()
- acro.probit() , acro.probitr()











Using the same api as pandas, adding checks for cell count and dominance

- acro.crosstab()
- acro.pivot_table()

Using the same api as statsmodels, adding checks for DoF:

- acro.logit(), acro.logitr()
- acro.ols(), acro.olsr()
- acro.probit() , acro.probitr()

R versions are prefixed by "acro_"











Using the same api as pandas, adding checks for cell count and dominance

- acro.crosstab()
- acro.pivot_table()

Using the same api as statsmodels, adding checks for DoF:

- acro.logit(), acro.logitr()
- acro.ols(), acro.olsr()
- acro.probit() , acro.probitr()

R versions are prefixed by "acro_"

Also support R's built in 'table' command











Using the same api as pandas, adding checks for cell count and dominance

- acro.crosstab()
- acro.pivot_table()

Using the same api as statsmodels, adding checks for DoF:

- acro.logit(), acro.logitr()
- acro.ols(), acro.olsr()
- acro.probit() , acro.probitr()

R versions are prefixed by "acro_"

- Also support R's built in 'table' command
- Stata versions prefixed by "acro"
 - Code currently captures and translates commands: table, regress, probit, logit





























acro = ACRO()









- acro = ACRO()
- acro.suppress= [True, False]









- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()









- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)









- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)









- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)
- acro.add_exception(key, text)









- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)
- acro.add_exception(key, text)
- acro.add_comment(key, text)









- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)
- acro.add_exception(key, text)
- acro.add_comment(key, text)
- acro.custom_output(filename, description)











- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)
- acro.add_exception(key, text)
- acro.add_comment(key, text)
- acro.custom_output(filename, description)
- acro.finalise(output_directory,format)









Example: what happens if researcher requests a disclosive table

```
acro.suppress = False
safe_table = acro.crosstab(df.recommend, df.parents)
print(safe_table)
INFO:acro:get_summary(): fail; threshold: 4 cells may need suppressing;
INFO:acro:outcome_df:
parents
          |great_pret
                       |pretentious |usual
recommend |
not_recom | ok |
priority |
                    ok I
recommend | threshold; | threshold; | threshold;
spec prior |
very_recom | threshold; |
INFO:acro:records:add(): output_1
           great_pret pretentious usual
parents
recommend
not_recom
                 1440
                             1440
                                   1440
                  858
                                   1924
priority
                             1484
recommend
spec_prior
                                     758
                             1264
                 2022
                             132
                                     196
very_recom
```









SACRO Outputs	Viewer				Release and download		
output_1	crosstab table	► ACRO risk profile					
output_3	pivot_table table	7 None Hak prome					
output_4	ols regression						
output_5	olsr regression	output_1					
output_6	probit regression	Overstand at					
output_7	logit regression	Created at	21 August 2023 at 12:16				
pivot_table output_8	crosstab table custom	Output type	crosstab table				
e output_o	custom	ACRO status	Fail (threshold: 4 cells may need suppressing)				
		Comments	Please let me have this data.				
			6 cells were supressed in this table				
		Exception request	because I really want it				
		Review	You cannot approve this output until you add a comment. Approve Reject Review comments on output_1:				
		output_1_0.csv			Open file		
		output_i_o.csv			Орен не		
		recommend	great_pret	pretentious	usual		
		not_recom	1440	1440	1440		
		priority	858	1484	1924		
		recommend	0	0	2		
		spec_prior	2022	1264	758		
		very_recom	0	132	196		









	SACRO Outputs	Viewer				Release and downlo	
	output_1 output_3	crosstab table pivot_table table	► ACRO risk profile				
	output_4 cutput_5	ols regression olsr regression	output_1				
	output_6 cutput_1	probit regression logit regression	Created at	21 August 2023 at 12:16			
	pivot_table output_8	crosstab table custom	Output type	crosstab table			
utputs	. outpuco	custom	ACRO status	Fail (threshold: 4 cells ma	ay need suppressing)		
review			Comments	Please let me have this de6 cells were supressed in			
			Exception request	because I really want it			
			Review	You cannot approve this output until you add a comment. Approve Reject			
				Review comments on outp	ut_1:		
			output_1_0.csv			Open file	
			recommend	great_pret	pretentious	usual	
			not_recom	1440	1440	1440	
			priority	858	1484	1924	
			recommend	0	0	2	
			spec_prior	2022	1264	758	
			very_recom	0	132	196	









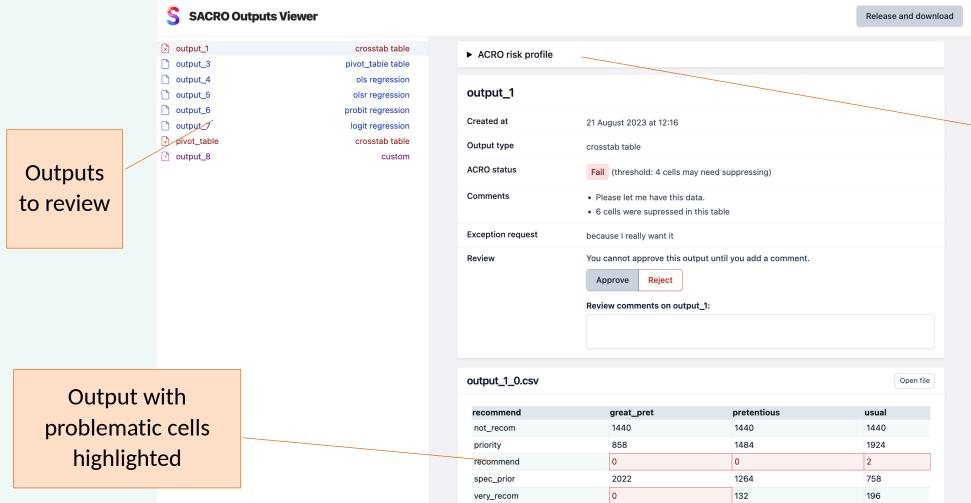
Outputs to review Output with problematic cells highlighted PACRO risk profile ACRO risk profile ACRO risk profile Output_1 Created at 21 August 2023 at 12:16 Output_1 Created at 21 August 2023 at 12:16 Output_1 Created at 21 August 2023 at 12:16 Output ype crosstab table Coutput ype crosstab table Coutput ype crosstab table Comments Please let me have this data. 6 cells were supressed in this table Exception request Dutput_1 Output_1 Output_1 Created at 21 August 2023 at 12:16 Output ype crosstab table Comments Please let me have this data. 6 cells were supressed in this table Exception request Documents Please let me have this data. 7 Comments Please let me have this data. 8 Review Comments on output_1: Output_1_O.csy Output_1_1 O		SACRO O	utputs Viewer				Release and download	
Outputs to review Outputs Please let me have this data. Gells were supressed in this table Exception request Decause I really want it Review You cannot approve this output until you add a comment. Approve Reject Review comments on output_1: Output with problematic cells highlighted Outputs Output with Please let me have this data. Output until you add a comment. Output to pretentious Issued Iss		output_3 output_4 output_5 output_6 output_7		► ACRO risk profile				
Outputs to review Comments Comment			olsr regression	output_1				
Outputs to review Comments Output with problematic cells highlighted Coutput storeview Comments Coutput storeview Comments Comments Comments Please let me have this data. Comments Please let me have this data. Comments Please let me have this data. Comments of cells were supressed in this table Exception request Decause I really want it Review You cannot approve this output until you add a comment. Approve Reject Review comments on output_1: Commend Great_pret Fecommend Great_pret Freetntious Base Jase J	() o		· · · · · ·	Created at	21 August 2023 at 12:16			
Outputs to review ACRO status Fail (threshold: 4 cells may need suppressing) Comments Please let me have this data. 6 cells were supressed in this table Exception request because I really want it Review You cannot approve this output until you add a comment. Approve Reject Review comments on output_1: Output with problematic cells highlighted Tecommend Tecom				Output type	crosstab table			
Exception request because I really want it Review You cannot approve this output until you add a comment. Approve Reject Review comments on output_1: Output with problematic cells highlighted Tecommend great_pret pretentious usual not_recom 1440 1440 1440 priority 858 1484 1924 highlighted Tecommend 0 0 0 2 spec_prior 2022 1264 758			custom	ACRO status	Fail (threshold: 4 cells may need suppressing)			
Review You cannot approve this output until you add a comment. Approve Reject Review comments on output_1: Output with problematic cells highlighted Pecommend great_pret pretentious usual not_recom 1440 1440 1440 1440 1440 priority 858 1484 1924 1924 1440 1440 1440 1440 1440 1440 1440 14	to review			Comments				
Output with problematic cells highlighted Output with priority 858 1484 1924 1924 1926 1926 1926 1926 1926 1926 1926 1926				Exception request	because I really want it			
Output with problematic cells highlighted Problematic cells spec_prior Output with precommend great_pret pretentious usual not_recom 1440 1440 1440 1440 1440 1440 1440 144				Review	Approve Reject			
problematic cells highlighted recommend great_pret pretentious usual 1440 1440 1440 priority 858 1484 1924 recommend 0 0 0 2 spec_prior 2022 1264 758	Output wi	th		output_1_0.csv			Open file	
highlighted 858 1484 1924 recommend 0 0 2 spec_prior 2022 1264 758				recommend	great_pret	pretentious	usual	
highlighted 858 1484 1924 recommend 0 0 2 spec_prior 2022 1264 758	problematic	cells		not_recom	1440	1440	1440	
spec_prior 2022 1264 758	highlighte	tad				-		
	Highlighte	:u						











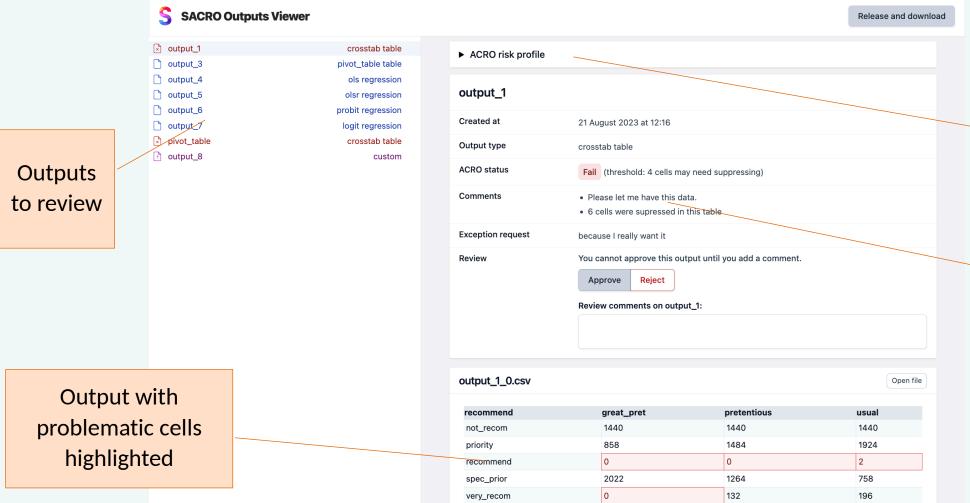
view TRE risk appetite











view TRE risk appetite

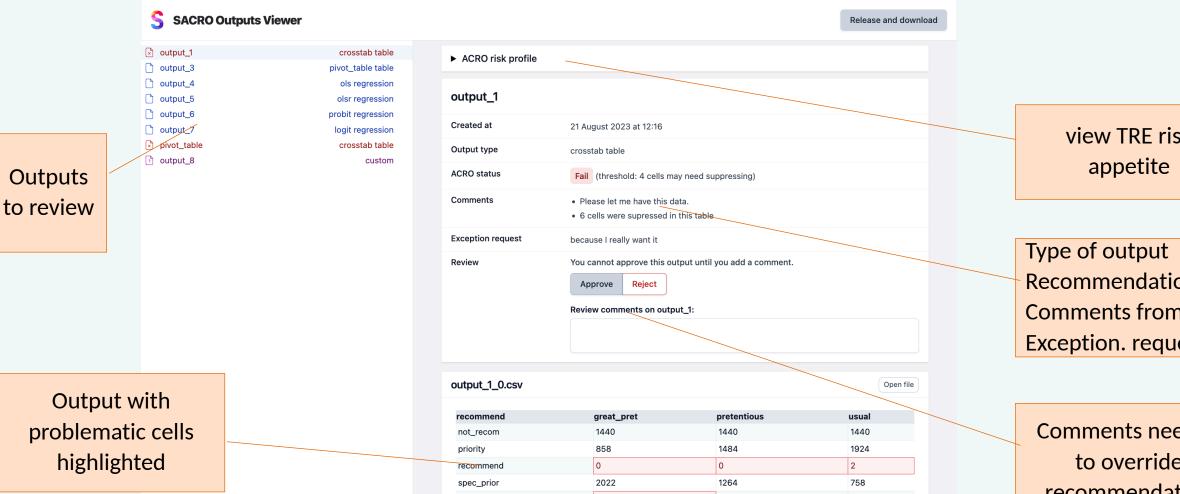
Type of output **Recommendation** Comments from user Exception. request











view TRE risk

Recommendation Comments from user Exception. request

Comments needed to override recommendation

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2020

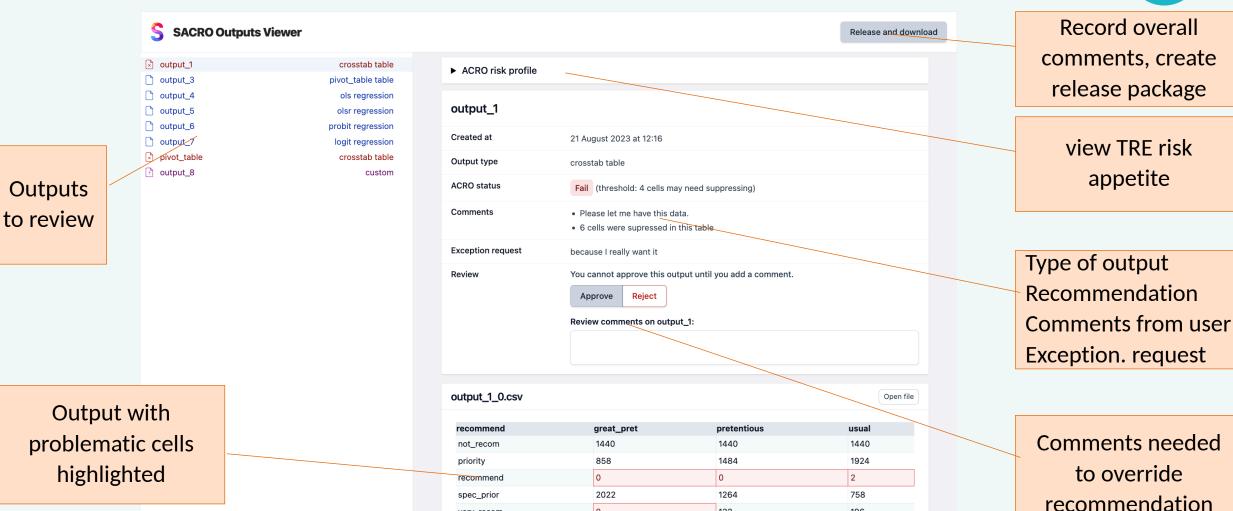
very_recom











SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2020

very_recom











Get involved Thanks to:









Get involved Thanks to:

ACRO 'engine':

• https://github.com/AI-SDC/ACRO











Get involved

Thanks to:

ACRO 'engine':

• https://github.com/AI-SDC/ACRO

Viewer:

• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk









Get Involved

Get involved Thanks to:

ACRO 'engine':

• https://github.com/AI-SDC/ACRO

Viewer:

https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk











Get involved

ACRO 'engine':

• https://github.com/AI-SDC/ACRO

Viewer:

• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

 All the project partners and TRE staff for all their feedback on the tools as they developed











Get involved

ACRO 'engine':

• https://github.com/AI-SDC/ACRO

Viewer:

• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement











Get involved

ACRO 'engine':

• https://github.com/AI-SDC/ACRO

Viewer:

• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group











Get involved

ACRO 'engine':

• https://github.com/AI-SDC/ACRO

Viewer:

• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group
- You for listening











Get involved

ACRO 'engine':

• https://github.com/AI-SDC/ACRO

Viewer:

• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group
- You for listening











Get involved

ACRO 'engine':

• https://github.com/AI-SDC/ACRO

Viewer:

• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group
- You for listening











Get involved

ACRO 'engine':

• https://github.com/AI-SDC/ACRO

Viewer:

• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group
- You for listening

This work is funded by UK research and Innovation, [Grant Number MC_PC_23006], as part of Phase 1 of the DARE UK (Data and Analytics Research Environments UK) programme, delivered in partnership with Health Data Research UK (HDR UK) and Administrative Data Research UK (ADR UK)